日　本　国　特　許　庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００１年　７月　４日

出願番号
Application Number: 特願２００１－２０２９１８

出願人
Applicant(s): 株式会社日立製作所

２００１年　８月１７日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造

出証番号　出証特２００１－３０７２５８

【書類名】　特許願
【整理番号】　K01000481A
【あて先】　特許庁長官殿
【国際特許分類】　G06F　3/06
【発明者】
【住所又は居所】　神奈川県小田原市中里３２２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
【氏名】　吉田　　晃
【特許出願人】
【識別番号】　000005108
【氏名又は名称】　株式会社日立製作所
【代理人】
【識別番号】　100075096
【弁理士】
【氏名又は名称】　作田　康夫
【手数料の表示】
【予納台帳番号】　013088
【納付金額】　21,000円
【提出物件の目録】
【物件名】　明細書　1
【物件名】　図面　1
【物件名】　要約書　1
【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ディスクアレイ制御装置における共有メモリコピー機能

【特許請求の範囲】

【請求項１】

ホストコンピュータとのインターフェースを有するチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報および前記磁気ディスク装置の管理情報を格納する共有メモリ部と，

前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と，前記チャネルインターフェース部および前記ディスクインターフェース部と前記共有メモリ部を接続する手段と，前記各部を駆動する電源供給手段を有し、前記ホストコンピュータからのデータのリード／ライト要求に対し，前記チャネルインターフェース部は、

前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより，データのリード／ライトを行うディスクアレイ制御ユニットを，複数ユニット有するディスクアレイ制御装置であって，前記複数のディスクアレイ制御ユニット内の前記共有メモリ部間を接続する手段と，前記複数のディスクアレイ制御ユニット内の前記キャッシュメモリ部間を接続する手段を有し，前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置。

【請求項２】

ホストコンピュータとのインターフェースを有するチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格

納するキャッシュメモリ部と、前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報および前記磁気ディスク装置の管理情報を格納する共有メモリ部と，

前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と，前記チャネルインターフェース部および前記ディスクインターフェース部と前記共有メモリ部を接続する手段と，前記各部を駆動する電源供給手段を有し、前記ホストコンピュータからのデータのリード／ライト要求に対し，前記チャネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、

前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより，データのリード／ライトを行うディスクアレイ制御ユニットを，複数ユニット有するディスクアレイ制御装置であって，前記複数のディスクアレイ制御ユニット内の前記共有メモリ部間を接続する手段と，前記複数のディスクアレイ制御ユニット内の前記キャッシュメモリ部間を接続する手段を有し，該接続手段を介して，前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であり、コピー処理中においても、ディスクアレイ制御ユニット内の前記チャネルインターフェース部および前記ディスクインターフェース部から、前記コピー処理の領域に対する前記共有メモリ部格納データのリード／ライト処理可能であることを特徴とするディスクアレイ制御装置。

【請求項３】

前記複数のディスクアレイ制御ユニット内の前記複数のチャネルインターフェース部および前記複数のディスクインターフェース部と前記複数のキャッシュメモリ部との間は，前記複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続され、前記複数のチャネルインターフェース部および前記複数のディスクインターフェース部と前記複数の共有メモリ部との間は，前記複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網に

よって接続され、前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置。

【請求項４】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送手段により格納データのどの領域まで転送が終了しているかを記録する手段を持つことを特徴とする請求項１または２に記載のディスクアレイ制御装置。

【請求項５】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送手段は前記相互結合網の使用頻度をモニタする機能を持ち、前記相互結合網の使用頻度が低い時に前記転送を行う機能を持つことを特徴とする請求項１または２に記載のディスクアレイ制御装置。

【請求項６】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送中に転送先の前記共有メモリ部に読み出し要求があった場合、転送終了領域であるか否かを判断し、転送終了領域であれば前記転送先共有メモリ部から格納データを読み出し、転送終了未の領域であれば前記転送元共有メモリ部から格納データを読み出し、

前記転送中に転送先の前記共有メモリ部に書込み要求があった場合、前記転送先共有メモリ部および前記転送元共有メモリ部の対応領域に格納データを書き込む機能を持つことにより、前記転送中の他処理に影響を及ぼす事無く、非同期に転送を行う事が出来る機能を持つ事を特徴とする請求項１または２に記載のディスクアレイ制御装置。

【請求項７】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記

ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送中にアプリケーションプログラムは転送処理中であることを意識せずにリード／ライト処理可能である事を特徴とする請求項１または２に記載のディスクアレイ制御装置。

【請求項８】

前記ディスクアレイ制御ユニット内の前記共有メモリ部内に通常動作に使用しないリザーブ領域を有し、前記ディスクアレイ制御ユニット内の前記共有メモリ部内の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部内のリザーブ領域に前記相互結合網を介して転送する手段を持つことを特徴とする請求項１または２に記載のディスクアレイ制御装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、ディスクアレイ制御装置に関し、特に、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置技術に関する。また、複数のサーバおよびパソコンを接続した装置に関しても、本発明は利用可能である。

【０００２】

【従来の技術】

今般のコンピュータシステムにおいて、処理性能の向上に対する期待は大きく、特にディスクサブシステムのＩ／Ｏ性能向上に対する要求は高い。磁気ディスクを記憶媒体とするディスクサブシステム（以下「サブシステム」という。）のＩ／Ｏ性能は半導体記憶装置を記憶媒体とするコンピュータの主記憶のＩ／Ｏ性能に比べて、３～４桁程度小さく、従来からこの差を縮めること、すなわちサブシステムのＩ／Ｏ性能を向上させる努力がなされている。

【０００３】

また、銀行、証券、電話会社等に代表される大企業では、従来各所に分散していたコンピュータおよびストレージを、データセンターの中に集中化してコンピュータシステムおよびストレージシステム構成することにより、コンピュータシステムおよびストレージシステムの運用、保守、管理に要する費用を削減する傾

向にあり、特に大型／ハイエンドのストレージシステムには、数百台以上のホストコンピュータへ接続するためのチャネルインターフェースのサポート（コネクティビティ）、数百テラバイト以上の記憶容量のサポートが要求されている。

【０００４】

一方、近年のオープン市場の拡大、今後予想されるストレージ・エリア・ネットワーク（ＳＡＮ）の普及により、大型／ハイエンドのストレージシステムと同様の高機能・高信頼性を備えた小規模構成（小型筐体）のストレージシステムへの要求が高まっている。

【０００５】

サブシステムのＩ／Ｏ性能を向上させるための１つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクアレイと呼ばれるシステムが知られている。ディスクアレイの場合、上位コンピュータからのＩ／Ｏを記録する複数の磁気ディスク装置と、上位コンピュータのＩ／Ｏを受付け複数の磁気ディスク装置へ転送するディスクアレイ制御装置から構成されるのが一般的である。この中で、大規模接続・大容量への要求に対しては、従来の大型／ハイエンドのディスクアレイ制御装置を複数接続して超大規模なディスクアレイ制御装置を構成する方法が考えられる。ディスクアレイ制御装置にはディスクアレイ制御装置に関する制御情報（例えば、ディスクアレイ制御装置内のキャッシュメモリの管理情報等）を格納する共有メモリを保持することが知られている。

【０００６】

複数のディスクアレイ制御装置を接続することにより、共有メモリが複数のディスクアレイ制御装置に分散することになる。共有メモリが複数のディスクアレイ制御装置に分散することにより共有メモリ領域のコピーに伴う転送処理性能が、従来の一つのディスクアレイ制御装置によりシステムが構築されていた場合に比べ、ディスクアレイシステムの性能に与える影響が大きくなる。

【０００７】

例えば，従来技術では，図２に示すようにホストコンピュータ５０とディスクアレイ制御装置２との間のデータ転送を実行する複数のチャネルＩＦ部１１と、

磁気ディスク装置５とディスクアレイ制御装置２間のデータ転送を実行する複数のディスクＩＦ部１２と、磁気ディスク装置５のデータを一時的に格納するキャッシュメモリ部１４と，ディスクアレイ制御装置２に関する制御情報（例えば、チャネルＩＦ部１１およびディスクＩＦ部１２とキャッシュメモリ部１４との間のデータ転送制御に関する情報，磁気ディスク装置５に格納するデータの管理情報）を格納する共有メモリ部１３とを備え、１つのディスクアレイ制御装置２内において，共有メモリ部１３およびキャッシュメモリ部１４は全てのチャネルＩＦ部１１およびディスクＩＦ部１２からアクセス可能な構成となっている。

【０００８】

このディスクアレイ制御装置２では、チャネルＩＦ部１１およびディスクＩＦ部１２と共有メモリ部１３との間、およびチャネルＩＦ部１１およびディスクＩＦ部１２とキャッシュメモリ部１４との間は，相互結合網２３，および相互結合網２４でそれぞれ接続される。

【０００９】

チャネルＩＦ部１１は、ホストコンピュータ５０と接続するためのインターフェースおよびホストコンピュータ５０に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクＩＦ部１２は、磁気ディスク装置５と接続するためのインターフェースおよび磁気ディスク装置５に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクＩＦ部１２は、RAID機能の実行も行う。

【００１０】

この従来のディスクアレイ制御装置２では，ディスクアレイ制御装置内に共有メモリ部１３が存在する為、共有メモリ間で共有する情報を持つ必要が無く、共有メモリ間でコピーが必要になった場合についても共有メモリ部が同一装置内にある為、転送に伴う相互接続網の競合等の他アクセスに対する影響が小さい。

【００１１】

また，コピー機能について、米国特許５，６８０，６４０に開示されている従来技術では，旧記憶装置から新記憶装置へとデータを移行する際にオンラインで実行する方法が示されている。ここでは、新記憶装置内に、旧記憶装置内ボリュ

ームの各アドレス（トラック）毎にテーブルを有し、旧記憶装置から新記憶装置へデータ移行が完了したかどうかをトラック毎に記憶する。移行中にホストからのＩ／Ｏ要求があった場合には、そのテーブルを参照して動作を決定する。例えばリード要求があった場合、テーブルを参照して、そのリード要求のあったレコード（ブロック）が新記憶装置に移行されているかどうかをチェックし、旧記憶装置からのデータが移行されていない時には旧記憶装置からデータを読み込む。

【００１２】

新記憶装置内にデータがあれば新記憶装置内からデータを読み出す。また、書込み要求があった場合には、新記憶装置にデータ書込みを行い、テーブルの更新を行う。本方式は、記憶装置を置き換える場合のコピー機能について示されているが、本方式を複数のディスクアレイ制御装置により構成されるシステムに応用することは可能である。

【００１３】

【発明が解決しようとする課題】

米国特許５，６８０，６４０に開示されている従来技術では，コピー元の記憶装置内全てのラックについてテーブルを保持しなければならない。通常テーブルには半導体メモリが使用され

コピー元とコピー先各々にテーブルを持つことにより、コストが高くなる。

【００１４】

また、この方法では、データ移行中のホストからの書込みでコピー先にのみデータを書き込む為、移行中にどちらかの記憶装置に障害が発生した場合、どちらの記憶装置もデータに矛盾のある状態となってしまう。また、この方式は、記憶装置の置き換え時のデータコピーについての技術であり、動作中の複数のディスクアレイ制御装置間のデータコピーについて記述したものではない。

【００１５】

【課題を解決するための手段】

上記課題およびディスクアレイ制御装置間の共有メモリ間コピー機能を実現するため、本発明におけるディスクアレイ制御装置は、下記の構成を取る。目的は、ホストコンピュータとのインターフェースを有するチャネルインターフェース

部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャネルインターフェース部およびディスクインターフェース部とキャッシュメモリ部との間のデータ転送に関する制御情報および磁気ディスク装置の管理情報を格納する共有メモリ部と，チャネルインターフェース部およびディスクインターフェース部とキャッシュメモリ部を接続する手段と，チャネルインターフェース部およびディスクインターフェース部と共有メモリ部を接続する手段と，各部を駆動する電源供給手段を有し、

ホストコンピュータからのデータのリード／ライト要求に対し，チャネルインターフェース部は、ホストコンピュータとのインターフェースとキャッシュメモリ部との間のデータ転送を実行し、ディスクインターフェース部は、磁気ディスク装置とキャッシュメモリ部との間のデータ転送を実行することにより，データのリード／ライトを行うディスクアレイ制御ユニットを，複数ユニット有するディスクアレイ制御装置であって，複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する手段と，複数のディスクアレイ制御ユニット内のキャッシュメモリ部間を接続する手段を有し，共有メモリ部間を接続する手段とキャッシュメモリ部間を接続する手段は，

おのおの独立に動作し，相互に影響を与えない手段を有し，ディスクアレイ制御ユニット内のチャネルインターフェース部およびディスクインターフェース部から，他のディスクアレイ制御ユニット内の前記共有メモリ部のデータ，またはキャッシュメモリ部のデータをリード／ライト可能であり、ディスクアレイ制御ユニット内の共有メモリ部と他のディスクアレイ制御ユニット内の共有メモリ部間で格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置により達成される。

【００１６】

好ましくは，複数のディスクアレイ制御ユニット内の複数のチャネルインターフェース部および複数のディスクインターフェース部と複数のキャッシュメモリ部との間は，複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続し、複数のチャネルインターフェース部および複数のディス

クインターフェース部と複数の共有メモリ部との間は，複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続する。

【００１７】

共有メモリの内容を、他ディスクアレイ制御ユニット内の他共有メモリにコピーする際に、複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網を介してデータが移行される。

【００１８】

コピーに先立ち、コピー開始アドレスとコピー終了アドレスおよびコピー開始情報をコピー先共有メモリ部および共有メモリ間に跨るスイッチ部の情報レジスタに登録を行う。このコピー開始情報の登録により、コピー元のマイクロプログラムは、コピーが終了したものとし、

次の処理に移行可能となる。ハードは、コピー開始情報の登録完了により、以降のコピー先共有メモリへの読み出しアクセス発生時、コピー情報レジスタとアクセスアドレスの比較により該アドレスがコピー完了未領域か否かの情報を得、コピー完了未領域へのアクセスの場合は、

コピー元共有メモリへのアクセスに切り替えコピー元共有メモリから該当データを読み出しリクエスト発行元にデータを送出する。また、リクエストがコピー完了済み領域へのアクセスであった場合コピー先共有メモリからデータを読み出し、リクエスト発行元にデータを送出する。リクエスト発行元のマイクロプログラムは、該当エリアがコピー中であることを意識する必要は無い。

【００１９】

その他、本願が開示する課題、およびその解決方法は、発明の実施形態の欄および図面により明らかにする。

【００２０】

【発明の実施形態】

以下、本発明の実施例を図面を用いて説明する。

［実施例１］

図１および図３に、本発明の一実施例を示す。

【００２１】

図１に示すように，ディスクアレイ制御装置１は複数のディスクアレイ制御ユニット１－２から構成される。ディスクアレイ制御ユニット１－２は、ホストコンピュータ５０とのインターフェース部（チャネルＩＦ部）１１と、磁気ディスク装置５とのインターフェース部（ディスクＩＦ部）１２と、共有メモリ部１３と、キャッシュメモリ部１４を有し，チャネルＩＦ部１１およびディスクＩＦ部１２と共有メモリ部１３の間は、ディスクアレイ制御ユニット１－２内において直接に接続されている。また、複数のディスクアレイ制御ユニット１－２の間では、共有メモリ部１３は相互結合網２２を介して接続され、キャッシュメモリ部１４は相互結合網２１を介して接続されている。

【００２２】

すなわち、相互結合網２１、あるいは相互結合網２２を介して、全てのチャネルＩＦ部１１およびディスクＩＦ部１２から、全ての共有メモリ部１３、あるいは全てのキャッシュメモリ部１４へアクセス可能な構成となっている。共有メモリ部１３内には、コピー開始情報および転送情報を格納するコピー情報レジスタ１５と共有メモリ１６を持つ。

【００２３】

図１にて，ディスクアレイ制御ユニット１－２間の共有メモリ領域のコピー処理について考えると、チャネルＩＦ部１１内のマイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、チャネルＩＦ部１１よりコピー元の共有メモリ領域１７が存在する共有メモリ部１３内のコピー情報レジスタ１５にコピー元共有メモリアドレスおよびコピー先共有メモリアドレスとコピー開始を示すバリッドビットをセットする。同様に、コピー先の共有メモリ領域１８が存在する他ディスクアレイ制御ユニット１－２内の共有メモリ部１３内のコピー情報レジスタ１５にもコピー情報をセットする。

【００２４】

コピー情報レジスタ１５へのコピー情報登録完了をコピー先共有メモリ部１３からコピー要求元のチャネルＩＦ部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。コピー元共有メモリ領域１７が存在する共有メモリ部１３は、コピー情報レジスタのコピ

ー実行ビットがバリッドとなったことにより、共有メモリ１６内のコピー元領域１７の値を読み出し、相互結合網２４を介し、他ディスクアレイ制御ユニット１－２内の共有メモリ部１３に転送する。転送データ到着により、コピー先共有メモリ領域１８を保持する他ディスクアレイ制御ユニット１－２内共有メモリ部１３は、転送データをコピー先領域１８に書き込む。

【００２５】

同様な処理をコピー領域全てについて行い、全ての領域のコピー終了により、各々の共有メモリ部１３内のコピー情報レジスタ１５内のコピー実行ビットのバリッドをＯＦＦにする。コピー実行中を示すコピー情報レジスタ１５内のコピー実行ビットがバリッドの場合、コピー先共有メモリ部１３内のコピー先領域１８へのコピー処理に伴うアクセス以外のアクセスがあった場合、コピー先共有メモリ部１３内の共有メモリ１６にアクセスするリクエストは全て、コピー情報レジスタ１５内のコピー終了未領域アドレスとアドレス比較することにより、アクセス領域がコピー終了未領域であった場合は、該当アクセスが読み出し要求の場合、ハードによりコピー元共有メモリ部１３にリクエストを転送し、コピー元領域１７からデータを読み出し、アクセス要求元にデータ転送を行う。

【００２６】

該当アクセスが、書込みアクセスの場合は、コピー先領域１８の該当アドレス部とコピー元領域１７の該当アドレス部に書込みを行う。アクセス領域がコピー終了領域およびコピー対象領域以外の場合は、コピー先共有メモリ１６の該当領域に対し、読み出しおよび書込み動作を行う。

【００２７】

図３によりディスクアレイ制御ユニット１－２間の共有メモリ領域のコピー処理におけるコピー先の共有メモリ部１３内の処理について考えると、マイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、ディスクアレイ制御ユニット１－２間を跨る相互結合網２２を介し、共有メモリアクセスパス２００により、コピー先の共有メモリ領域１８が存在する共有メモリ部１３内のコピー情報レジスタ１５内のコピー元開始アドレス１０２、コピー先開始アドレス１０３およびコピー先終了アドレス１０４にコピー情報をセットする。更に、

コピー先の共有メモリ領域１８が存在する共有メモリ部１３内のコピー情報レジスタ１５内のコピー実行ビット１０１にコピー開始を示すバリッドビットをセットする。

【００２８】

コピー実行ビット１０１がセットされていることにより、コピー情報レジスタ１５内のアドレス生成論理１０７、コピー領域判定論理１０８およびコピー終了判定論理１０９が機能することになる。コピー情報レジスタ１５へのコピー情報登録完了を共有メモリ部１３からコピー要求元のチャネルＩＦ部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。

【００２９】

もちろん、コピー先共有メモリ１６へのコピー領域全ての書込み終了を待って、他処理に移行する方式としても良い。コピー先共有メモリ部１３は、コピー実行ビット１０１がセットされていることにより、コピー処理に伴うコピーリクエストに対し、共有メモリ内のコピー先領域１８に対し、コピー情報を書き込む。コピー処理に伴うコピーリクエスト受付により、コピー情報レジスタ１５内のカウンタ１０５を更新し、コピー先開始アドレス１０３と加算することにより、コピー実行アドレス１０６に現在のコピー実行アドレスをセットする。

【００３０】

コピーリクエストを受け付けるたびに、カウンタ１０５をカウントアップし、コピー実行アドレス１０６を更新する。コピー先終了アドレス１０４とコピー実行アドレス１０６の値をコピー終了判定論理１０９により比較することにより、コピー終了を判定する。コピー終了判定により、コピー実行ビット１０１をリセットし、コピー処理終了となる。コピー元共有メモリ部にも同様なコピー終了判定論理を持つことにより、コピー元共有メモリ部にコピー終了を知らせる必要は無くなる。

【００３１】

もちろん、コピー終了をコピー元共有メモリ部およびマイクロプログラムに知らせる方式としても良い。次に、コピー処理中の、コピー処理以外の共有メモリ

１６へのリクエストに対する処理について考えると、コピー処理中のコピー処理以外の共有メモリ１６へのリクエストに対しては、共有メモリ部１３内でコピー処理以外の共有メモリ１６へのリクエストを優先とした方が性能的に有利である。コピー処理以外の共有メモリ１６へのリクエスト受付により、共有メモリ部１３は、コピー情報レジスタ１５内のコピー実行ビット１０１がセットされている場合、コピー実行中と判断し、コピー領域判定論理１０８により、コピー実行アドレス１０６とコピー先終了アドレス１０４の値とリクエストアドレスの値を比較することにより、リクエストアドレスがコピー終了未領域へのアクセスか否かを判定する。コピー領域判定論理１０８により、リクエストアドレスがコピー終了未領域へのアクセスと判定した場合、アドレス生成論理１０７によりコピー元開始アドレス１０２の値とコピー先開始アドレス１０３の差分およびリクエストアドレスとの加算により、コピー元共有メモリへのアクセスアドレスを求め、コピー元共有メモリ部に他共有メモリアクセスパス２０２により、相互結合網２２を介し、リクエストを転送する。

【００３２】

書込みリクエストの場合、コピー終了未領域への書込みでも、自共有メモリ１６の該当アドレスにも書き込む方式を取った方が、障害および性能面で有利である。コピー終了未領域への書込みの場合、更新済である情報を共有メモリ部１３内に持つことにより（図示せず）、以降の該当領域のリクエストに対しても自共有メモリ１６内へのリクエストとして受け付けることが可能となり、コピー元領域を含む他ディスクアレイ制御ユニット内のコピー元共有メモリ部にリクエストを転送する必要が無くなる。

【００３３】

コピーまた、コピー領域判定論理１０８により、リクエストアドレスがコピー終了領域およびコピー対象領域以外へのアクセスと判定した場合、アドレス生成論理１０７によりリクエストアドレスを選択し、自共有メモリアクセスパス２０１を介し、自共有メモリ１６内にアクセスする。

［実施例２］

図４に、本発明の他の実施例を示す。

図４に示すように、複数のディスクアレイ制御ユニット１－２からなるディスクアレイ制御装置１の構成は、複数のディスクアレイ制御ユニット１－２を跨ぐ、チャネルＩＦ部１１およびディスクＩＦ部１２とキャッシュメモリ部１４間の接続構成および共有メモリ部１３間の接続構成を除いて、実施例１の図１に示す構成と同様である。図４に示すように、ディスクアレイ制御ユニット１－２内のキャッシュメモリ部１４と他ディスクアレイ制御ユニット１－２内のキャッシュメモリ部１４は、筐体間キャッシュスイッチ１４０を介して、筐体間キャッシュメモリパス１４１により相互に接続されている。

【００３４】

障害を考え筐体間キャッシュスイッチ１４０は、冗長性を持たせ二重化構造となっている。また、同様に筐体間キャッシュメモリパス１４１についても二重化構成となっている。筐体間キャッシュメモリパス１４１および筐体間キャッシュスイッチ１４０を介すことにより、複数のディスクアレイ制御ユニット１－２間で他ディスクアレイ制御ユニット１－２内のキャッシュメモリ部１４にアクセスすることが可能となる。

【００３５】

一般にキャッシュおよびディスクへアクセスするパスのデータ転送は、数キロバイトの転送が発生する為大規模であるのに対し、共有メモリをアクセスするパスのデータ転送は、データが制御情報の為、数バイト程度で数サイクルの規模である。相互結合網が一つの場合、共有メモリをアクセスするパスとキャッシュおよびディスクへアクセスするパスが同じ相互結合網を使用することになり、小規模アクセスである共有メモリアクセスの性能低下が生じることになる。このため、図４に示すように、ディスクアレイ制御ユニット１－２内の共有メモリ部１３と他ディスクアレイ制御ユニット１－２内共有メモリ部１３は、筐体間キャッシュメモリスイッチ１４０とは別の筐体間共有メモリスイッチ１３０を介して、筐体間共有メモリパス１３１により相互に接続されている。

【００３６】

障害を考え筐体間共有メモリスイッチ１３０は、冗長性を持たせ二重化構造となっている。また、同様に筐体間共有メモリパス１３１についても二重化構成と

なっている。筐体間キャッシュメモリスイッチ１４０および筐体間キャッシュメモリパス１４１と筐体間共有メモリスイッチ１３０および筐体間共有メモリパス１３１は、各々独立に動作可能である。共有メモリ部１３内のコピー情報レジスタ１５内の構造は、図３にて示してあるコピー情報レジスタ１５の構造と同一である。

【００３７】

また、図４に示すように、筐体間共有メモリスイッチ１３０内にも同様にコピー情報レジスタ１５を持つ構造となっている。　図４によりディスクアレイ制御ユニット１－２間の共有メモリ領域のコピー処理における筐体間共有メモリスイッチ１３０内の処理について考えると、マイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、ディスクアレイ制御ユニット１－２間を跨る筐体間共有メモリパス１３１および筐体間共有メモリスイッチ１３０を介し、コピー先の共有メモリ領域１８が存在する共有メモリ部１３内のコピー情報レジスタ１５内のコピー元開始アドレス、コピー先開始アドレスおよびコピー先終了アドレスにコピー情報をセットする。

【００３８】

更に、コピー先の共有メモリ領域１８が存在する共有メモリ部１３内のコピー情報レジスタ１５内のコピー実行ビットにコピー開始を示すバリッドビットをセットする。同様に、筐体間共有メモリスイッチ１３０内のコピー情報レジスタ１５内のコピー元開始アドレス１０２、コピー先開始アドレス１０３およびコピー先終了アドレス１０４にコピー情報をセットする。更に、筐体間共有メモリスイッチ１３０内のコピー情報レジスタ１５内のコピー実行ビット１０１にコピー開始を示すバリッドビットをセットする。コピー実行ビット１０１がセットされていることにより、コピー情報レジスタ１５内のアドレス生成論理１０７、コピー領域判定論理１０８およびコピー終了判定論理１０９が機能することになる。

【００３９】

コピー先の共有メモリ領域１８が存在する共有メモリ部１３内および筐体間共有メモリスイッチ１３０内のコピー情報レジスタ１５へのコピー情報登録完了をコピー先の共有メモリ領域１８が存在する共有メモリ部１３内および筐体間共有

メモリスイッチ１３０からコピー要求元のチャネルＩＦ部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。

【００４０】

もちろん、コピー先共有メモリ１６へのコピー領域全ての書込み終了を待って、他処理に移行する方式としても良い。コピー先共有メモリ部１３は、コピー実行ビットがセットされていることにより、コピー処理に伴うコピーリクエストに対し、共有メモリ内のコピー先領域１８に対し、コピー情報を書き込む。筐体間共有メモリスイッチ１３０は、コピー処理に伴うコピーリクエスト受付により、コピー情報レジスタ１５内のカウンタ１０５を更新し、コピー先開始アドレス１０３と加算することにより、コピー実行アドレス１０６に現在のコピー実行アドレスをセットする。

【００４１】

コピーリクエストを受け付けるたびに、カウンタ１０５をカウントアップし、コピー実行アドレス１０６を更新する。コピー先終了アドレス１０４とコピー実行アドレス１０６の値をコピー終了判定論理１０９により比較することにより、コピー終了を判定する。コピー終了判定により、コピー実行ビット１０１をリセットし、コピー処理終了となる。コピー元共有メモリ部にも同様なコピー終了判定論理を持つことにより、コピー元共有メモリ部にコピー終了を知らせる必要は無くなる。もちろん、コピー終了を筐体間共有メモリスイッチ１３０およびコピー先共有メモリ部１３より、コピー元共有メモリ部およびマイクロプログラムに知らせる方式としても良い。次に、コピー処理中の、コピー処理以外の共有メモリ１６へのリクエストに対する処理について考えると、コピー処理中のコピー処理以外の共有メモリ１６へのリクエストに対しては、筐体間共有メモリスイッチ１３０内でコピー処理以外の共有メモリ１６へのリクエストを優先とした方が性能的に有利である。

【００４２】

コピー処理以外の共有メモリ１６へのリクエスト受付により、筐体間共有メモリスイッチ１３０は、コピー情報レジスタ１５内のコピー実行ビット１０１がセ

ットされている場合、コピー実行中と判断し、コピー領域判定論理１０８により、コピー実行アドレス１０６とコピー先終了アドレス１０４の値とリクエストアドレスの値を比較することにより、リクエストアドレスがコピー終了未領域へのアクセスか否かを判定する。コピー領域判定論理１０８により、リクエストアドレスがコピー終了未領域へのアクセスと判定した場合、アドレス生成論理１０７によりコピー元開始アドレス１０２の値とコピー先開始アドレス１０３の差分およびリクエストアドレスとの加算により、コピー元共有メモリへのアクセスアドレスを求め、コピー元共有メモリ部１３に筐体間共有メモリパス１３１により、リクエストを転送する。

【００４３】

また、コピー領域判定論理１０８により、リクエストアドレスがコピー終了領域およびコピー対象領域以外へのアクセスと判定した場合、アドレス生成論理１０７によりリクエストアドレスを選択し、筐体間共有メモリパス１３１を介し、該当共有メモリ領域を含むディスクアレイ制御ユニット１－２内の共有メモリ部１３にアクセスする。筐体間共有メモリスイッチ１３０にコピー情報レジスタ１５が無い場合には、コピー先領域１８へのコピー処理以外のリクエストが発生した場合に、アクセス元のディスクアレイ制御ユニット１－２から筐体間共有メモリパス１３１および筐体間共有メモリスイッチ１３０を介し、コピー先領域１８を含む他ディスクアレイ制御ユニット１－２内の共有メモリ部１３にアクセスした後、コピー未領域と判定され、再び筐体間共有メモリパス１３１および筐体間共有メモリスイッチ１３０を介して、コピー元領域１７を含む他ディスクアレイ制御ユニット１－２内の共有メモリ部１３にアクセスする必要があるのに対し、筐体間共有メモリスイッチ１３０にコピー情報レジスタ１５を持つことにより、コピー先領域１８を含む他ディスクアレイ制御ユニット１－２内の共有メモリ部１３にアクセスせずにアクセス領域がコピー未領域か否かの判定を得ることができ、処理の高速化および筐体間共有メモリパス１３１の専有時間を削減でき性能面で有利である。

【００４４】

本実施例によれば、共有メモリ部間でコピー処理を行う場合、コピー実行元の

マイクロプログラムは、コピー情報をコピー情報レジスタ１５に登録することにより、コピー処理終了となり、次処理に移行することが可能となる。コピー情報レジスタ１５へのコピー情報登録以降は、コピー中であることを意識せずに全ての共有メモリへのアクセスが可能となる。コピー対象領域へのアクセスについても、アドレスの変更およびリクエストの振り分けをマイクロプログラムは意識する必要が無い。

【００４５】

また、コピー処理動作より通常処理を優先とする論理を組み込むことにより、コピー処理に伴う、他処理への性能面の影響を少なくし、また、キャッシュアクセスの筐体間接続網と共有メモリの筐体間接続網を別々に持つことにより、共有メモリ部のコピー処理により、キャッシュアクセスに影響を及ぼすことがなくなる。また、共有メモリ部間の相互結合網を介しての転送手段において相互結合網の使用頻度をモニタする機能を持たせ、相互結合網の使用頻度が低い時に共有メモリ部格納データの転送を行うことによりサブシステム性能の低下を抑制可能である。

［実施例３］

図５に、本発明の他の実施例を示す。

図５に示すように、共有メモリ部１３内の共有メモリ１６内の構成を除いて、実施例１の図１に示す構成と同様である。図５に示すように、各々のディスクアレイ制御ユニット１－２内の共有メモリ部１３内の共有メモリ１６内にリザーブ領域１９を持つことにより、ディスクアレイ制御ユニット１－２間の共有メモリ１６の再構成および再配置が可能となる。チャネルＩＦ部１１およびディスクＩＦ部１２とキャッシュメモリ１４との間のデータ転送に関する制御情報および磁気ディスク装置５の管理情報を格納する共有メモリ部１３は、各々自ディスクアレイ制御ユニット１－２内の情報を保持するほうが性能的に有利である。共有メモリ部１３が複数のディスクアレイ制御ユニット１－２に分散されて実装された場合、自ディスクアレイ制御ユニット１－２内の情報を常に自ディスクアレイ制御ユニット１－２内の共有メモリ部１３に保持することは難しいと考えられる。

【００４６】

図５により、動作中の共有メモリ再配置について考えると、モニタリング機構等（図示せず）により、自ディスクアレイ制御ユニット１－２内の情報が他ディスクアレイ制御ユニット１－２内の共有メモリ部１３に格納されていることを検出した場合、マイクロプログラムは自ディスクアレイ制御ユニット１－２内の情報が格納されている他ディスクアレイ制御ユニット１－２内の共有メモリ部１３内のコピー情報レジスタ１５にコピー元共有メモリアドレスおよびリザーブ領域アドレスとコピー開始を示すバリッドビットをセットする。

【００４７】

同様に、自ディスクアレイ制御ユニット１－２内の共有メモリ部１３内のコピー情報レジスタ１５にコピー元共有メモリアドレスおよびリザーブ領域アドレスとコピー開始を示すバリッドビットをセットする。コピー情報レジスタ１５へのコピー情報登録完了をコピー元共有メモリ部１３から再配置要求元のチャネルＩＦ部のマイクロプログラムに知らせることにより、マイクロプログラムは再配置処理完了とし他処理に移行可能となる。コピー元領域１７が存在する共有メモリ部１３は、コピー情報レジスタのコピー実行ビットがバリッドとなったことにより、共有メモリ１６内のコピー元領域１７の値を読み出し、相互結合網２４を介し、他ディスクアレイ制御ユニット１－２内の再配置要求元の共有メモリ部１３に転送する。転送データ到着により、再配置要求元共有メモリ部１３は、転送データをリザーブ領域１９に書き込む。同様な処理をコピー領域全てについて行い、全ての領域のコピー終了により、各々の共有メモリ部１３内のコピー情報レジスタ１５内のコピー実行ビットのバリッドをＯＦＦにする。

【００４８】

コピー実行中を示すコピー情報レジスタ１５内のコピー実行ビットがバリッドの場合、コピー元共有メモリ部１３内のコピー元領域１７へのコピー処理に伴うアクセス以外のアクセスがあった場合、コピー元共有メモリ部１３内の共有メモリ１６にアクセスするリクエストは全て、コピー情報レジスタ１５内のコピー終了未領域アドレスとアドレス比較することにより、アクセス領域がコピー終了未領域およびコピー対象領域以外の場合は、コピー元共有メモリ部１３内のコピー元共有メモリ１６の該当領域に対し、読み出しおよび書込み動作を行う。アクセ

ス領域がコピー終了領域の場合は、該当アクセスが読み出し要求の場合、コピー元共有メモリ部１３内のコピー元共有メモリ１６の該当領域からデータを読み出し、アクセス要求元にデータ転送を行い、該当アクセスが、書込みアクセスの場合は、コピー元領域１７の該当アドレス部と再配置要求元の共有メモリ部１３内のリザーブ領域１９の該当アドレス部に書込みを行う。

【００４９】

本実施例によれば、共有メモリ部間で再配置を行う場合、再配置要求元のマイクロプログラムは、コピー情報をコピー情報レジスタ１５に登録することにより、再配置処理終了となり、次処理に移行することが可能となる。コピー情報レジスタ１５へのコピー情報登録以降は、コピー中であることを意識せずに全ての共有メモリへのアクセスが可能となる。コピー対象領域へのアクセスについても、アドレスの変更およびリクエストの振り分けをマイクロプログラムは意識する必要が無い。また、コピー処理動作より通常処理を優先とする論理を組み込むことにより、コピー処理に伴う、他処理への性能面の影響を少なくし、また、キャッシュアクセスの筐体間接続網と共有メモリの筐体間接続網を別々に持つことにより、共有メモリ部のコピー処理により、キャッシュアクセスに影響を及ぼすことがなくなる。また、共有メモリ部間で、メモリ部格納データの転送中にアプリケーションプログラムは転送処理中であることを意識せずにリード／ライト処理可能である。

【００５０】

【発明の効果】

本発明によれば、複数台のディスクアレイ制御装置を１つのディスクアレイ制御装置として運用しようとする場合，複数のディスクアレイ制御装置に共有メモリが分散されて搭載された場合の、共有メモリ間でのコピー処理による性能低下を抑え，台数に比例した性能を出せるディスクアレイシステムを提供すること，また，ディスクアレイ制御装置が有する機能を，性能低下を抑えて複数台のディスクアレイ制御装置で実現することが可能となる。

【図面の簡単な説明】

【図１】

本発明によるディスクアレイ制御装置の構成を示す図。

【図２】

図１に示すディスクアレイ制御ユニット内の詳細構成を示す図。。

【図３】

従来のディスクアレイ制御装置の他の構成を示す図。

【図４】

本発明によるディスクアレイ制御装置の他の構成を示す図。

【図５】

本発明によるディスクアレイ制御装置の他の構成を示す図。

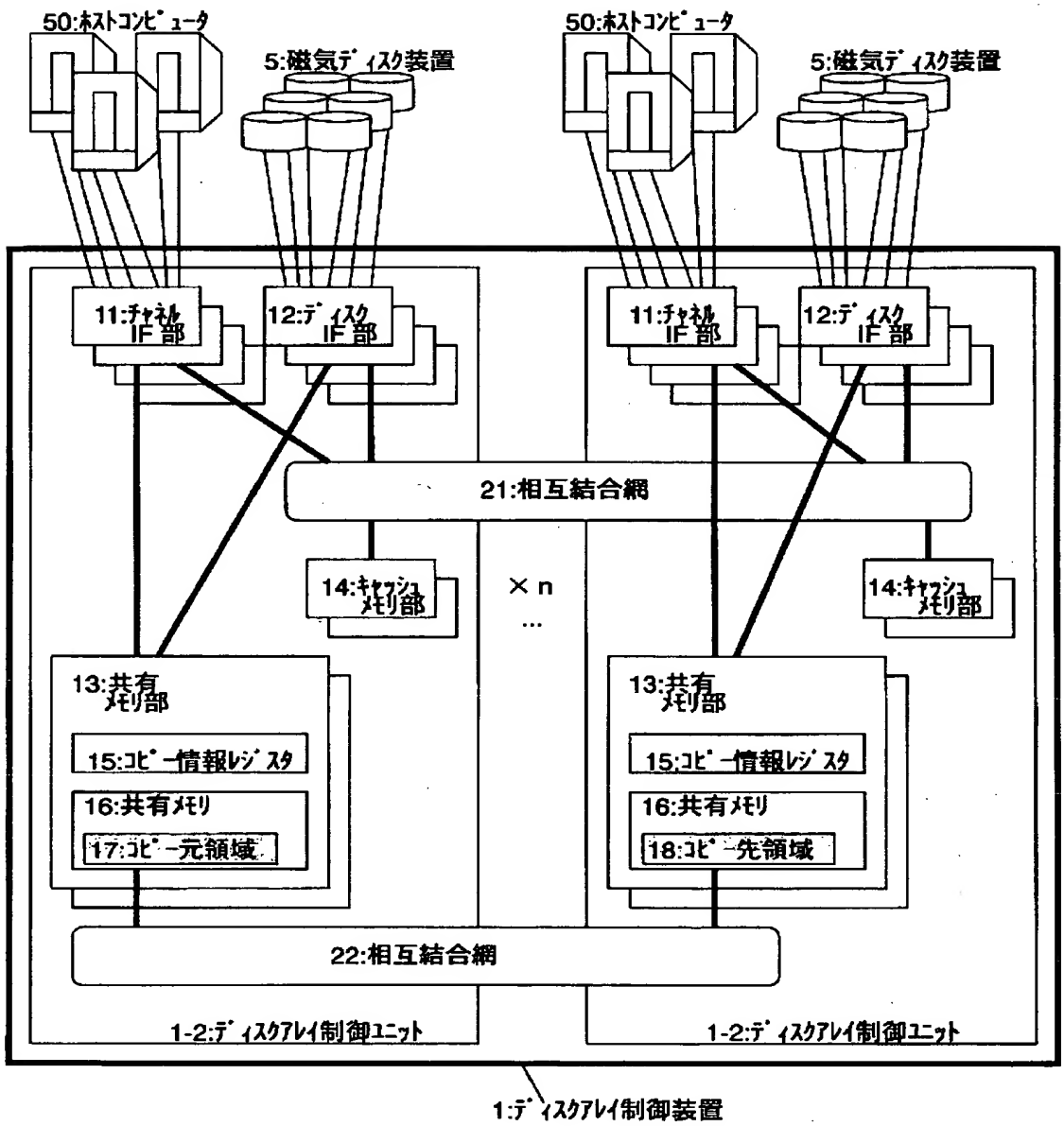
【符号の説明】

１，２：ディスクアレイ制御装置，１－２…ディスクアレイ制御ユニット、５…磁気ディスク装置，１１…チャネルＩＦ部、１２…ディスクＩＦ部、１３…共有メモリ部、１４…キャッシュメモリ部、１５…コピー情報レジスタ、２１，２２，２３，２４…相互結合網、５０…ホストコンピュータ

【書類名】　図面

【図１】

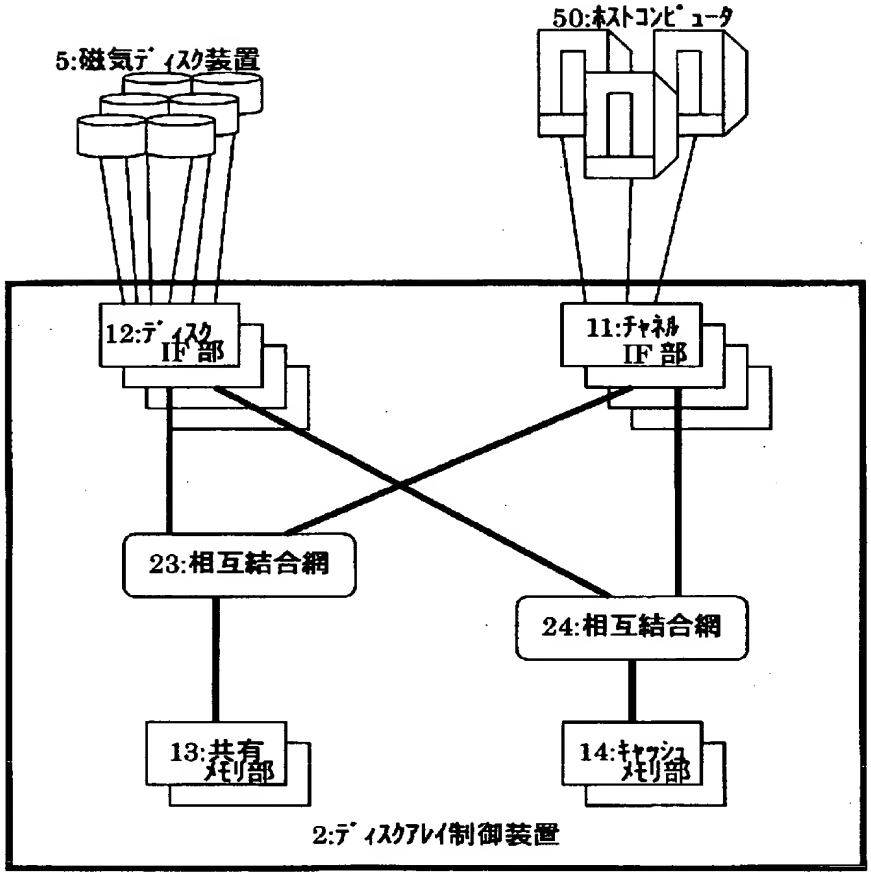
図１



出証特２００１－３０７２５８１

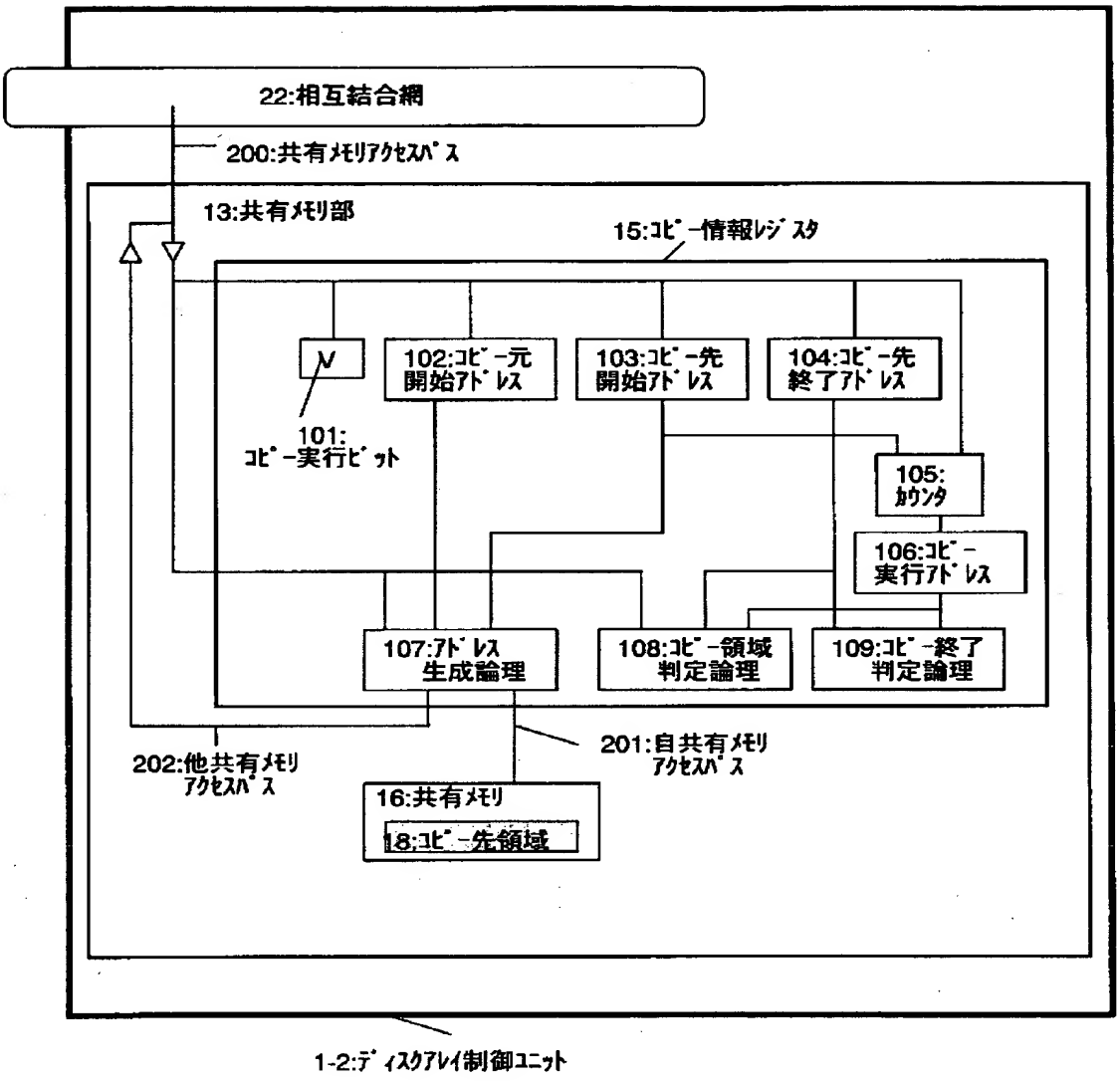
【図２】

図２


5:磁気ディスク装置
50:ホストコンピュータ
12:ディスクIF部
11:チャネルIF部
23:相互結合網
24:相互結合網
13:共有メモリ部
14:キャッシュメモリ部
2:ディスクアレイ制御装置

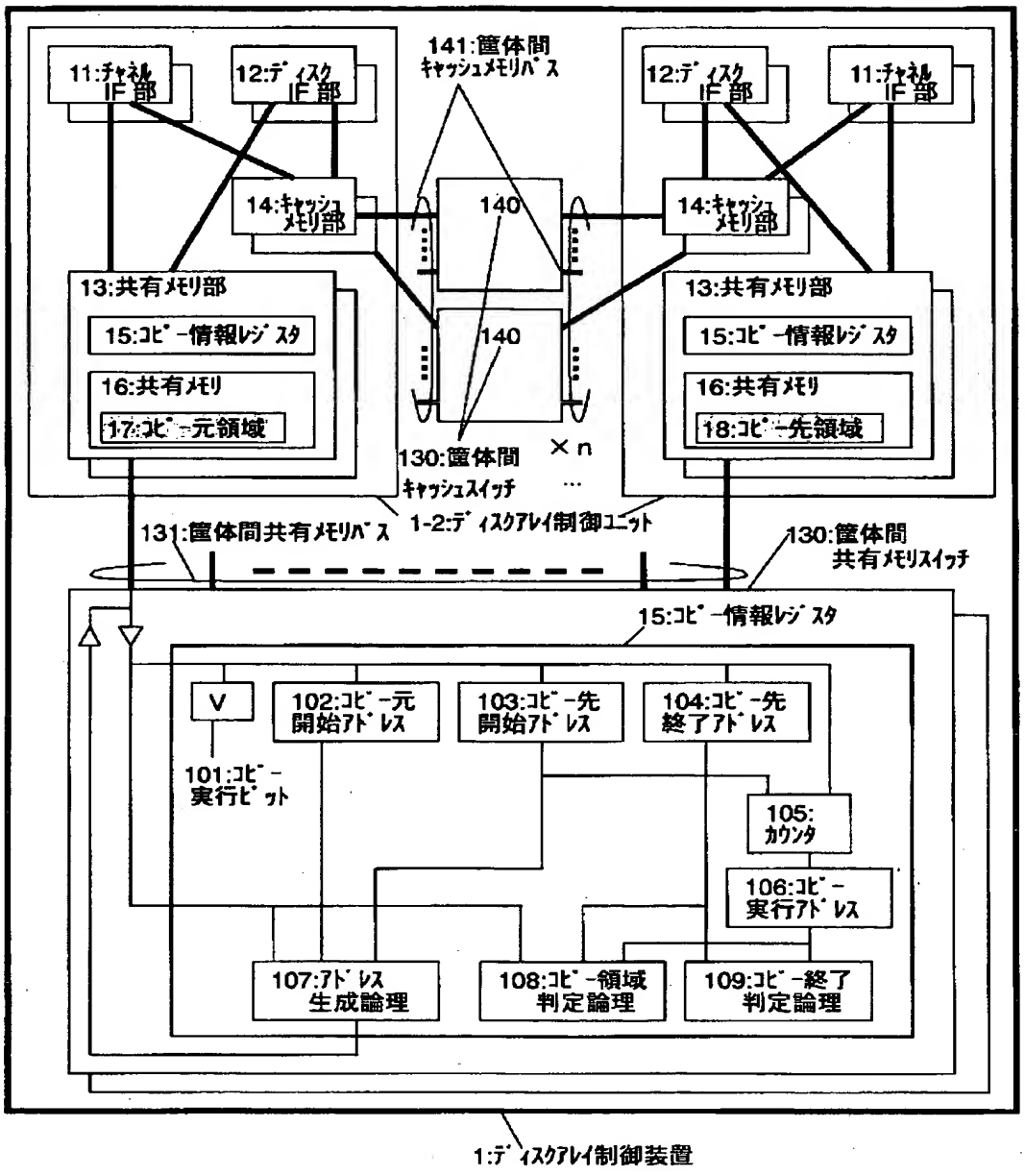
【図３】

図３


22:相互結合網
200:共有メモリアクセスパス
13:共有メモリ部
15:コピー情報レジスタ
V
101:
コピー実行ビット
102:コピー元
開始アドレス
103:コピー先
開始アドレス
104:コピー先
終了アドレス
105:
カウンタ
106:コピー
実行アドレス
107:アドレス
生成論理
108:コピー領域
判定論理
109:コピー終了
判定論理
202:他共有メモリ
アクセスパス
201:自共有メモリ
アクセスパス
16:共有メモリ
18:コピー先領域
1-2:ディスクアレイ制御ユニット

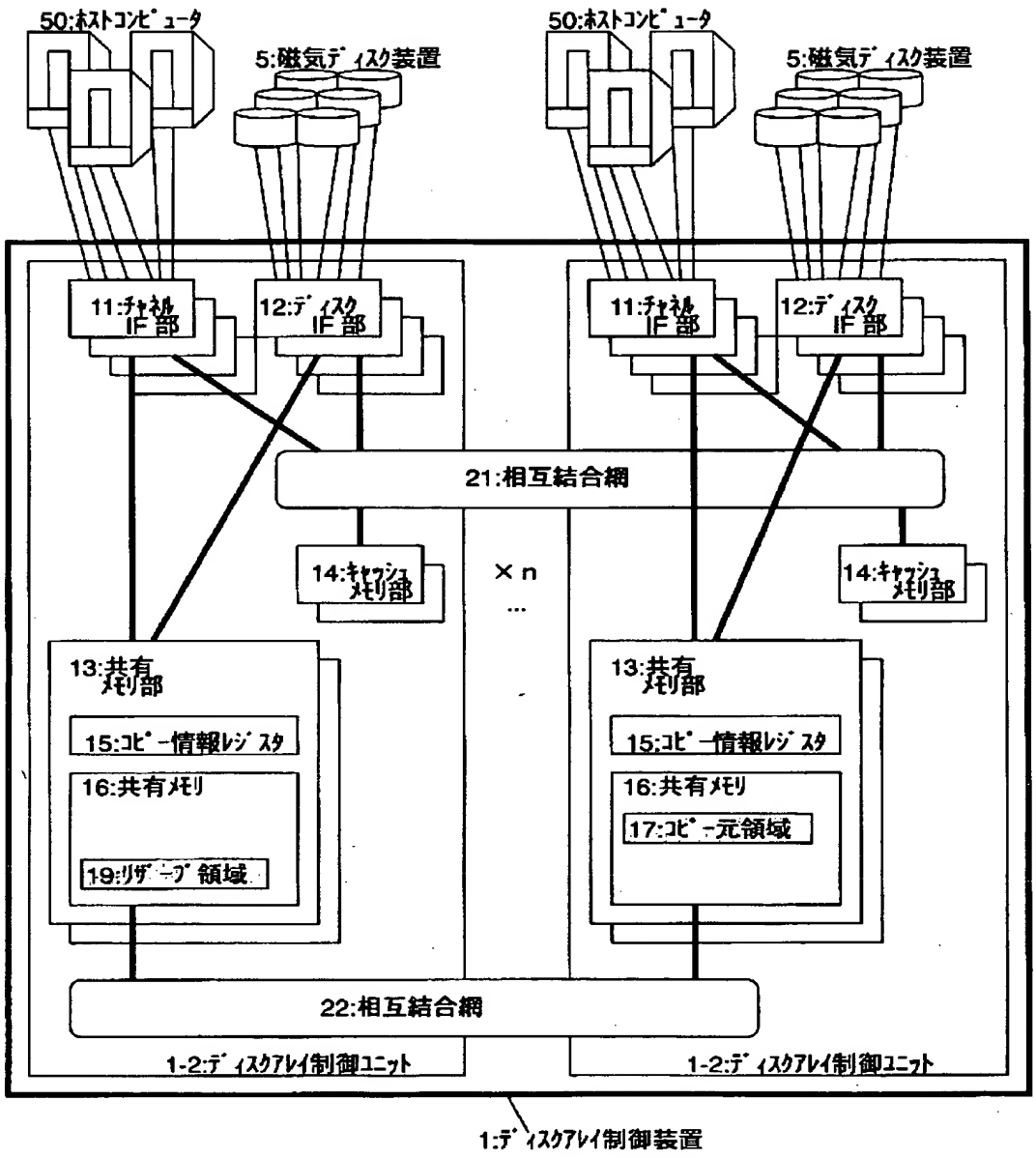
【図４】

図４


141:筐体間
ｷｬｯｼｭﾒﾓﾘﾊﾞｽ
11:ﾁｬﾈﾙ
IF部
12:ﾃﾞｨｽｸ
IF部
14:ｷｬｯｼｭ
ﾒﾓﾘ部
140
13:共有ﾒﾓﾘ部
15:ｺﾋﾟｰ情報ﾚｼﾞｽﾀ
16:共有ﾒﾓﾘ
17:ｺﾋﾟｰ元領域
18:ｺﾋﾟｰ先領域
130:筐体間
ｷｬｯｼｭｽｲｯﾁ
×n
1-2:ﾃﾞｨｽｸｱﾚｲ制御ﾕﾆｯﾄ
131:筐体間共有ﾒﾓﾘﾊﾞｽ
130:筐体間
共有ﾒﾓﾘｽｲｯﾁ
V
102:ｺﾋﾟｰ元
開始ｱﾄﾞﾚｽ
103:ｺﾋﾟｰ先
開始ｱﾄﾞﾚｽ
104:ｺﾋﾟｰ先
終了ｱﾄﾞﾚｽ
101:ｺﾋﾟｰ
実行ﾋﾞｯﾄ
105:
ｶｳﾝﾀ
106:ｺﾋﾟｰ
実行ｱﾄﾞﾚｽ
107:ｱﾄﾞﾚｽ
生成論理
108:ｺﾋﾟｰ領域
判定論理
109:ｺﾋﾟｰ終了
判定論理
1:ﾃﾞｨｽｸｱﾚｲ制御装置

【図５】

図５


50:ホストコンピュータ
5:磁気ディスク装置
50:ホストコンピュータ
5:磁気ディスク装置
11:チャネルIF部
12:ディスクIF部
11:チャネルIF部
12:ディスクIF部
21:相互結合網
14:キャッシュメモリ部
×n
14:キャッシュメモリ部
13:共有メモリ部
15:コピー情報レジスタ
16:共有メモリ
19:リザーブ領域
13:共有メモリ部
15:コピー情報レジスタ
16:共有メモリ
17:コピー元領域
22:相互結合網
1-2:ディスクアレイ制御ユニット
1-2:ディスクアレイ制御ユニット
1:ディスクアレイ制御装置

【書類名】要約書

【要約】

【課題】

複数台のディスクアレイ制御装置を１つのディスクアレイ制御装置として運用できるようにし、複数のディスクアレイ制御装置間での共有メモリ部のコピー処理による性能低下を抑え，台数に比例した性能を出せるディスクアレイシステムを提供することにある。

【解決手段】

上記課題は、チャネルＩＦ部と、ディスクＩＦ部と、キャッシュメモリ部と、共有メモリ部と，チャネルＩＦ部およびディスクＩＦ部とキャッシュメモリ部を接続する手段と，チャネルＩＦ部およびディスクＩＦ部と共有メモリ部を接続する手段と，前記各部を駆動する電源供給手段を有し、データのリード／ライトを行うディスクアレイ制御ユニットを，複数ユニット有するディスクアレイ制御装置であって，複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する手段と，複数のディスクアレイ制御ユニット内のキャッシュメモリ部間を接続する手段を有し、ディスクアレイ制御ユニット内にコピー情報レジスタを有し、また、複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する共有メモリスイッチ部にコピー情報レジスタを有することを特徴とするディスクアレイ制御装置により達成される。

【効果】

複数のディスクアレイ制御ユニット間でのコピー処理の性能の向上が可能で，ディスクアレイ制御ユニットの台数に比例してディスクアレイ制御装置の性能を向上できる

【選択図】　図１

認定・付加情報

特許出願の番号	特願２００１－２０２９１８
受付番号	５０１００９７５５３４
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１３年　７月　５日

＜認定情報・付加情報＞

【提出日】　平成13年 7月 4日

次頁無

出証特２００１－３０７２５８１

出　願　人　履　歴　情　報

識別番号　　〔０００００５１０８〕

１．変更年月日　　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所